The Computational Structure of Mental Representation

Joscha Bach

Berlin, February 26th, 2013

Now is the time for starting a new Artificial Intelligence initiative.

What is the mind?
–this is arguably the most interesting question that our species can ask itself.

Last month, a grand proposal by Barack Obama made headlines everywhere in the world: He suggested a large-scale initiative to create a detailed **map of the activity of the human brain**, at the level of individual neurons. This idea is quite similar to (and likely inspired by) Henry Markram's *Blue Brain* project in Lausanne, which recently won a 1.5Bn grant from the European Union. While this is an interesting project in its own right, it will not address the key question of cognitive science: What is the mind? What are the building blocks and fundamental operations of thinking and perception?

Here, I would like to suggest the instigation of a project that is at once more ambitious, more narrowly targeted, and likely to yield more profound theoretical, practical and cultural insights than the *Brain Activity Mapping* initiative: **The study of the computational structure of mental representation.**

Mental representation, not neurons will inform the core our understanding of cognition.

Cognition is incidentally enabled by human nervous systems, but in its nature, it is not a chemical, biological or physiological phenomenon. Instead, **cognitive systems** are a class of **information processing** systems, **thinking** is a set of certain, functionally identifiable **operations**, over certain, functionally identifiable types of **representations**, and with respect to a set of problems and properties given by certain **environments**.

Thus, the *Structure of Mental Representation* initiative should focus on mental content, with all its dynamic, relational, conceptual, and linguistic elements, and it's grounding in perception and interaction.

The study of mental representations requires a new scientific discipline

Unraveling mental representations will be an interdisciplinary effort among several cognitive sciences, but distinct from the disciplines that we find now:

The project will differ from linguistics, for instance, in similar way as geography differs from plate tectonics. Where linguists study the intrinsic 'geography' of languages, and the structural commonalities among them, mental representations uncover the underlying dynamics that produce natural languages (as the solutions to the problem of translating between hierarchical, distributed, associative, ambiguous representations and the discrete strings of symbols that we use as a means of exchanging and organizing ideas).

Integrating concepts from linguistics, cognitive psychology and neuroscience within the methodological framework of constructionist Artificial Intelligence

Mental representations lie outside the domain of neuroscience, which mostly focuses on material descriptions of the function of the underlying substrate. They cannot be studied well within contemporary psychology, which favors an experimentalist, quantitative approach, where we need to address qualitative questions by constructing working, implementable systems. And needless to say, the study of mental representations is currently not well represented in Artificial Intelligence research, which does provide a productive methodology, but has mostly turned towards applications and narrow AI solutions.

A new common ground between AI and cognitive science researchers

Despite the lack of a common methodological ground between the cognitive sciences, we can now observe a growing consensus on how to approach the problem of modeling the mind, and its representational apparatus. During the last decade, a number of initiatives have sprung up within AI, psychology and cognitive science, each with journals, workshops and conference series, and a large personal overlap among each community. Examples include Artificial General Intelligence (**AGI**), Biologically Inspired Cognitive Architectures (**BICA**), **Cognitive Systems**, and **Cognitive Modeling** (ICCM). Among those groups, there is an emerging consensus on several paradigms:

- **Integrated architectures of cognition**. We need to study whole, working systems, both for the purpose of comparison between approaches, but mainly, because cognition is not the product of the activity of individual modular functions, but of the interaction between them.

- **Universal representations**. Representations must provide both distributed and localist aspects, to enable neural learning, information retrieval via spreading activation, as well as symbolic processing (language and planning).

- **Perceptual grounding.** Representations should be grounded in an environmental interaction context, using a bottom-up/top-down perception paradigm, to allow for

open-ended autonomous learning and language acquisition.

- **Integration of perception and action.** Environmental interaction is (at least partially) a control process that requires a modeling of the relationship between sensory data and operations performed by the system.

- **(Semi-) Universal problem solving**, enabling paradigms of learning, planning, reasoning, analogies, language acquisition and reflection.

- A **structured memory**, including provisions for a world/situation model, a protocol of environmental interactions, procedural memory, declarative and typological abstractions, a model of self, and a 'mental stage', to facilitate anticipation and planning.

- **Decision making** and **motivational mechanisms**, to address both **autonomous**, goal-directed cognition in complex, open domains, and the genesis of goals and intentions.

- The direction and modulation of **attention**, and the emergence of **emotion and affect**, which are configurational aspects of cognition that either reflect the allocation of cognitive resources, or structure social interaction.

Constraints for Mental Representations

Based on the emergent paradigm of a new generation of cognitive architectures, we can think about mental representations in a new and productive way. More specifically, we know that mental representations include perceptual and propositional/conceptual content and we are aware of many constraints. Mental representations must

- Offer support for both **connectionist** and **symbolic** processing (including compositionality and grammatical structures), with the latter one being a special case of the former

- Possess a **hierarchical structure**, bottoming out in **sensory perception**

- Cover **prototypes**, **individuals** and **abstractions**

- Include **perceptual** and **relational features**, **objects**, **situations**, and **episodic knowledge**

- Allow simulation of **dynamic processes**

- Solve the **bridge problem** between fuzzy associative hierarchies and the discrete symbol strings of **natural language**

Operations on these representations include **reinforcement learning** and **classification**, perceptual **top-down/bottom-up** processing, **abstraction**, **analogy formation**, associative and syllogistic **reasoning**, **planning**, **reflection**, **anticipation** and several modes of **reorganization**.

Creating a scientific community

The task of the *Structure of Mental Representation* initiative will at first consist in the creation of a large and competitive community, built around a set of benchmark problems.

I suggest picking a set of problems that covers most of the above architectural requirements: the **comprehension of dynamic visual sequences** (movies) and **narratives**. The evaluation of comprehension may focus both on a **discourse level** (asking the system general and specific questions about the consumed visuals or narratives) and directly, by producing **dynamic renderings** and depictions of knowledge represented within the system. At least on the discourse level, a **direct comparison** to the functional properties of **child performance** at different cognitive stages is possible.

The Comprehension Challenge

Turning the benchmark task into a regular competition allows a direct comparison between models, and offers a strong incentive for exchange of solutions among research groups. The need to for broad solutions with given material and intellectual resources will enforce a higher degree of the reuse of code and ideas than we currently see in AI architectures (outside of robotic soccer, where such competitions have turned out to be highly successful).

By structuring the comprehension tasks into different levels and sub-domains (such as basic language acquisition tasks, grammar, perceptual and logical tasks, social reasoning/theory of mind, affective evaluation, constructive problem solving, use of analogies and metaphor etc.), we can formulate consistent long-term goals and realistic short-term problems.

The Formation of a Pilot Team

To pull the *Structure of Mental Representation* initiative off the ground, we will need to create a **pilot team**, which is the object of this proposal. The pilot team has the following tasks:

- Bring a **critically sized group** of people together, with a **well-defined** set of **shared goals**

- Formulate an **initial set of hypotheses**

- Bring these in the context of an existing or hypothetical cognitive architecture, so **specifications** can be derived

- Build an **initial software infrastructure** for testing benchmark problems, and to create a common methodological ground

Composition of a pilot team

The pilot team should consist of a **scientific board**, including a mutually compatible sub-set of contemporary thinkers in the field. Its core should consist of a new generation of researchers that have the resources and impetus to work full-time on the **design**, **implementation** and **publication** of the project's infrastructure. The initial team size should not be large, but **agile** and effective (I would suggest about **five permanent researchers,** about the same number of dedicated **software developers**, and 2–3 slots for **visiting scientists** at a time).

The initial duration should be about **2.5 years**, followed be an **evaluation**, and a possible continuation of similar length.

Possible candidates for the scientific board include (in no particular order) Paul Rosenbloom (who worked with Alan Newell on the cognitive architecture Soar, and now started a new architecture), John Sowa (well-known for his groundbreaking work in semantic networks), Ben Goertzel (designer of the cognitive architecture OpenCog), Luc Steels (an AI researcher studying the emergence of natural language in communities of software agents), Jerome Feldman (who devised building blocks for a *Neural Theory of Language*), Stephen Pinker (who thought much about the relationship between language and mental representation), Stephen Kosslyn (an expert on the psychology of mental representation), Daniel Dennet, Cristiano Castelfranchi (an AI researcher specializing on mental representations of intentions and social context), and many others.

The project should start out with a number of **kickoff meetings** to invite the interaction between suitable candidates, and the formation of a core group. The eventual pilot team could be located in many possible places, however, the project should be able to draw on the vicinity of academic and other infrastructure, and the influence of a creative environment.

About me (Joscha Bach):

Driven by an interest to learn hot the mind works, I studied philosophy and psychology, graduated in computer science/artificial intelligence, lead several small academic research groups, built the cognitive architecture *MicroPsi* and obtained a PhD in cognitive science. I co-founded a couple of startup companies in Berlin, but turned back towards academia to contribute to the question of how to build a mind.